On-Line Nonlinear Programming as a Generalized Equation

Mihai Anitescu

anitescu@mcs.anl.gov

Joint Work with: Victor Zavala

Motivation

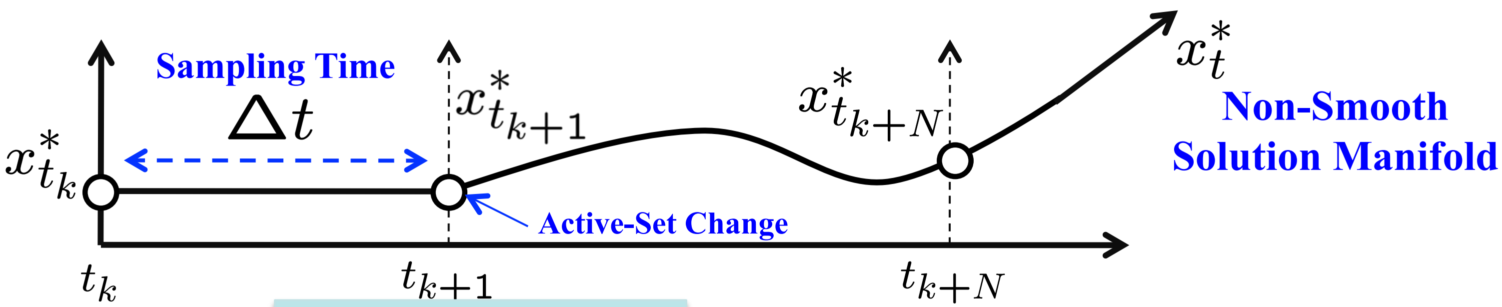
On-Line Optimization: MPC, MHE, RTO, Finance

Minutes → **Seconds**




Data Updated at Predefined Sampling Times

Decisions Obtained by Solving NLP/QP with Current Data



$$\min_{x \in X} \; f(x, t_k) \quad \text{s.t. } c(x, t_k) = 0$$

$$\boxed{\min_{x \in X} \; f(x, t_{k+1}) \quad \text{s.t. } c(x, t_{k+1}) = 0}$$

$$\min_{x \in X} \; f(x, t_{k+N}) \quad \text{s.t. } c(x, t_{k+N}) = 0$$

Objective: **Accommodate Large-Scale Dynamic Models in Suitable Time Scales**

Property: **Problems Close to Each Other! Can we exploit this to ensure stability?**

An abstract view of the issues

- **Rolling horizon optimal control:** $F(w,t)=0 \Rightarrow w=w(t)$ **the optimal control manifold.**
  - We already wrote the optimality conditions to get it here
  - F can be an operator that includes differential equations for dynamics, which can be discretized somehow.
  - w includes state variables, control variables and Lagrange multipliers
- **The variable** w **cannot be computed instantly, so we must allow it a time** Δt **.**
  - The problem becomes $F(w(t^k),t^k)=0; \quad F(w(t^{k+1}),t^{k+1})=0; \quad t^{k+1}=t^k+\Delta t$
- **Better, but we cannot guarantee that we find a solution in** Δt **even now. What if we solve the subproblem inexactly, e.g only its linearization or an inexact linearization?**

$$F(w^k,t^k)+\nabla_w F(w^k,t^k)(w^{k+1}-w^k)+\nabla_t F(w^k,t^k)\Delta t+r^k=0;$$

- **Could it work? Yes, if we can track the manifold (stability):**

$$\left\|w^k-w(t^k)\right\|\le O\left((\Delta t)^p\right)$$

- **Can we track the manifold with as little computation per time step as possible, particularly when inequality constraints are included (limited ramps, limited resources, sufficient supply) ? --- This becomes our central investigation issue.**
- **Can we do this in the limit of rapidly increasing information?** $\Delta t \to 0$

Outline of the Talk

1. **Nonlinear Programming**

2. **Generalized Equations**

- Single QP per Sampling Time

- Stability of NLP Error as $\Delta t \rightarrow 0$

3. **Augmented Lagrangean Strategy**

- Cheap Strategies for QP Solution – Projected Gauss Seidel

4. **Numerical Case Study**

5. **Conclusions and Future Work**

1. Nonlinear Programming

Nonlinear Programming

$$\begin{aligned} \min_x \ & f(x,t) \\ \text{s.t. } & c(x,t) = 0 \\ & x \geq 0 \end{aligned}$$

Active-Set SQP

$$\begin{aligned} \nabla_x f(x,t) + \nabla_x c(x,t)\,\lambda - \nu &= 0 \\ c(x,t) &= 0 \\ x^{(i)} &= 0,\ \forall i \in \mathcal{A} \\ \nu^{(i)} &= 0,\ \forall i \notin \mathcal{A} \end{aligned}$$

Interior Point

$$\begin{aligned} \nabla_x f(x,t) + \nabla_x c(x,t)\,\lambda - \nu &= 0 \\ c(x,t) &= 0 \\ X \cdot V &= \mu e \end{aligned}$$

QP Solver

$$\begin{aligned} \min \quad & \nabla_x f(x^k,t)^T \Delta x + \tfrac{1}{2}\Delta x^T \mathbf{H} \Delta x \\ \text{s.t.} \quad & c(x^k,t) + \mathbf{A}^T \Delta x = 0 \\ & \Delta x \geq -x^k \end{aligned}$$

KKT System

$$\begin{bmatrix} \mathbf{H} & \mathbf{A} & -I \\ \mathbf{A}^T & & \\ V & & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda \\ \Delta\nu \end{bmatrix} = - \begin{bmatrix} \mathbf{r}_x(w^k,t) \\ \mathbf{r}_\lambda(w^k,t) \\ X^k \cdot V^k - \mu e \end{bmatrix}$$

Newton Step Computation

$$\left[\begin{array}{ccc|cc} \mathbf{H} & \mathbf{A} & -I & & \\ \mathbf{A}^T & & & & \\ \mathbf{V} & & \mathbf{X} & \mathbf{E}_x & \mathbf{E}_\nu \\ \hline \mathbf{E}_x^T & & & & \\ & & \mathbf{E}_\nu^T & & \end{array}\right] \left[\begin{array}{c} \Delta x \\ \Delta\lambda \\ \Delta\nu \\ \hline \Delta\sigma_x \\ \Delta\sigma_\nu \end{array}\right] = - \left[\begin{array}{c} \mathbf{r}_x(w^k,t) \\ \mathbf{r}_\lambda(w^k,t) \\ \mathbf{r}_c(w^k,t) \\ \hline \mathbf{r}_{\sigma_x}(w^k,t) \\ \mathbf{r}_{\sigma_\nu}(w^k,t) \end{array}\right]$$

Interior-Point: - **Fixed Matrix Structure - No Symbolic Factorization Needed**

Active-Set: - **Changing Matrix Structure**
- **Each Internal QP Iteration is as Expensive as Outer IP Iteration**

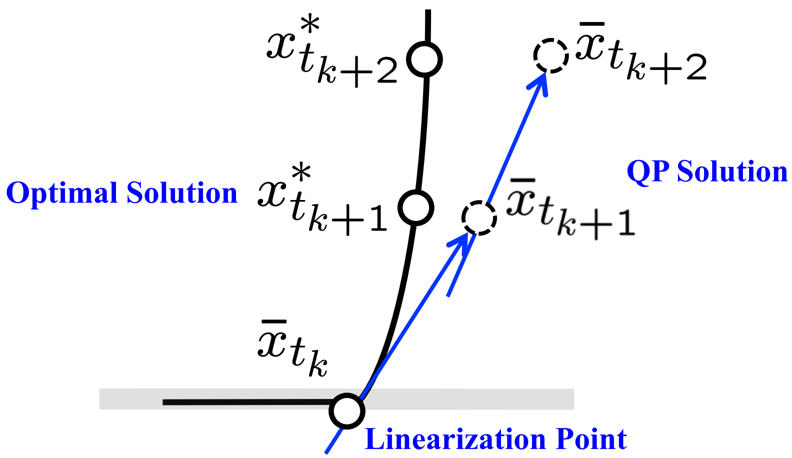
Newton Steps Accurate but Overhead is High. Limits attainable Δt !

Nonlinear Programming

A "Fast" NLP Solver is NOT Enough …

Approximate NLP Strategies

- <u>One</u> Quadratic Program (QP) Per Sampling Time *de Oliveira & Biegler, 1995, Diehl, et.al.,2001, Ohtsuka, 2004*
- Accurate But Slow vs. Approximate But Fast?
- The Dynamic System Escapes if we Insist in Accurate Solution …



Data

$$\begin{aligned} \min \quad & \nabla_x f(\bar{x}_{t_k}, t_{k+1})^T \Delta x + \tfrac{1}{2}\Delta x^T \nabla_{xx}\mathcal{L}(\bar{x}_{t_k}, \bar{\lambda}_{t_k}, t_k)\Delta x \\ \text{s.t.} \quad & c(\bar{x}_{t_k}, t_{k+1}) + \nabla_x c(\bar{x}_{t_k}, t_k)^T \Delta x = 0 \\ & \Delta x \geq -\bar{x}_{t_k} \end{aligned}$$

Time-Dependent QP

Issues:

- <u>Stability</u> of NLP Error, Changing Active Sets
- Solving the QP as Quickly as Possible (If $\Delta t \rightarrow 0$ Cheap Steps are Enough!)

1. Generalized Equations

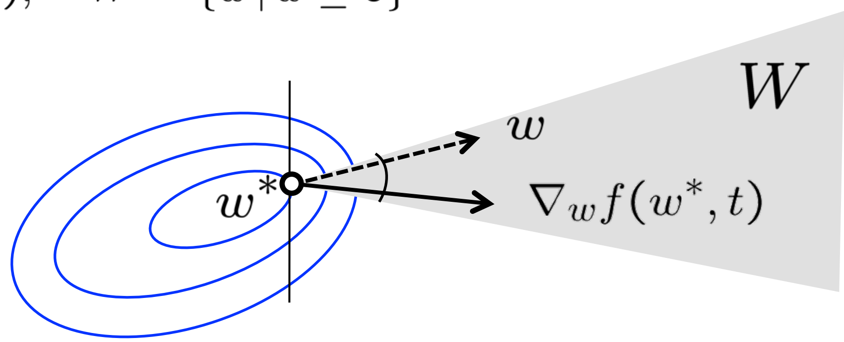
Generalized Equations

Generalized Equations (GE) *Robinson, 1977, 1980*

$-F(w,t) \in \mathcal{N}_W(w)$ $\leftarrow$ **Normal Cone Operator (compare with NLE)**

First-Order KKT Conditions of $\min\limits_{w \in W} \; f(w,t), \quad W = \{w \,|\, w \geq 0\}$

$$-\nabla_w f(w^*,t)^T(w^* - w) \geq 0, \; \forall w \in W$$



Canonical Linearized Generalized Equation (LGE)

$$\delta \in F(w^*_{t_0}, t_0) + \nabla_w F(w^*_{t_0}, t_0)(w - w^*_{t_0}) + \mathcal{N}_W(w) \qquad w(\delta) = \psi^{-1}[\delta]$$ $\leftarrow$ **Solution Operator**

Definition *(Robinson, 1977)*: **LGE is Strongly Regular at** $w^*_{t_0}$ **if** $\exists\, L_\psi \geq 0$ s.t. $\|w(\delta) - w^*_{t_0}\| \leq L_\psi \|\delta\|$

Theorem: ψ^{-1} **is Lipschitzian if:**

$$\mathrm{M} = \nabla_w F(w^*_{t_0}, t_0) = \begin{bmatrix} \mathrm{M}_{11} & \mathrm{M}_{12} & \mathrm{M}_{13} \\ \mathrm{M}_{21} & \mathrm{M}_{22} & \mathrm{M}_{23} \\ \mathrm{M}_{31} & \mathrm{M}_{32} & \mathrm{M}_{33} \end{bmatrix} \qquad \hat{\mathrm{M}} = \begin{bmatrix} \mathrm{M}_{11} & \mathrm{M}_{12} \\ \mathrm{M}_{21} & \mathrm{M}_{22} \end{bmatrix}$$

Active Degenerate Inactive

1. M_{11} **Non-Singular**
2. $\mathrm{M}_{22} - \mathrm{M}_{21}\mathrm{M}_{11}^{-1}\mathrm{M}_{12}$ **Is Positive Definite**

Generalized Equations

Context of NLP $\min_{x \in X} f(x,t), \text{ s.t. } c(x,t) = 0$

Solution of Perturbed LGE $\bar{w}_t = [\bar{x}_t \; \bar{\lambda}_t]$ **Around** $w^*_{t_0}$

$$0 \in F(w^*_{t_0}, t) + \nabla_w F(w^*_{t_0}, t_0)(w - w^*_{t_0}) + \mathcal{N}_W(w)$$

Canonical Form $\Updownarrow$

$$\delta \in F(w^*_{t_0}, t_0) + \nabla_w F(w^*_{t_0}, t_0)(w - w^*_{t_0}) + \mathcal{N}_W(w)$$

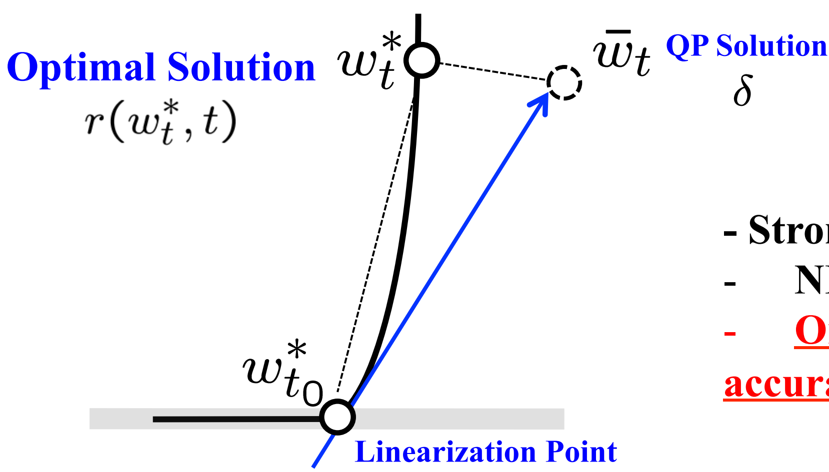
With $\delta = F(w^*_{t_0}, t_0) - F(w^*_{t_0}, t)$

KKT Conditions of Perturbed QP

$$\begin{array}{ll} \min & \nabla_x f(x^*_{t_0}, t)^T \Delta x + \frac{1}{2} \Delta x^T \nabla_{xx} \mathcal{L}(w^*_{t_0}, t_0) \Delta x \\ \text{s.t.} & c(x^*_{t_0}, t) + \nabla_x c(x^*_{t_0}, t_0)^T \Delta x = 0 \\ & \Delta x \geq -x^*_{t_0} \end{array}$$

From Lipschitz Continuity and Mean Value Theorem

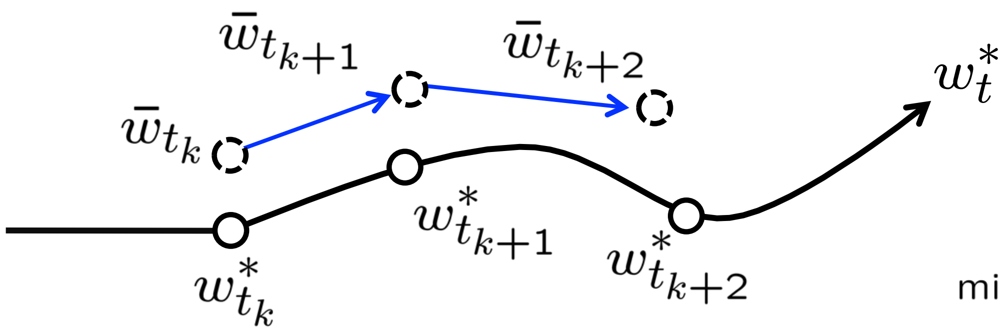
$$\begin{aligned} \|w^*_t - \bar{w}_t\| &\leq L_\psi \|r(w^*_t, t) - \delta\| \\ &\leq L_\psi \| \Big(F(w^*_{t_0}, t_0) + F_w(w^*_{t_0}, t_0)(w^*_t - w^*_{t_0}) - F(w^*_t, t) \Big) - \Big(F(w^*_{t_0}, t_0) - F(w^*_{t_0}, t) \Big) \| \\ &\leq L_\psi \| F_w(w^*_{t_0}, t_0)(w^*_t - w^*_{t_0}) - F(w^*_t, t) + F(w^*_{t_0}, t) \| \\ &\leq L \Delta t^2 \end{aligned}$$



- **Strong Regularity Requires SSOC and LICQ**
- **NLP Error is Bounded by LGE <u>Perturbation</u>**
- **<u>One QP solution from exact manifold is second-order accurat</u>**

Generalized Equations

But I am never EXACTLY on the manifold: Stability of uncentered NLP Error



Time-Dependent QP

$$
\begin{aligned}
\text{min} \quad & \nabla_x f(\bar{x}_{t_k}, t_{k+1})^T \Delta x + \tfrac{1}{2} \Delta x^T \nabla_{xx} \mathcal{L}(\bar{w}_{t_k}, t_k) \Delta x \\
\text{s.t.} \quad & c(\bar{x}_{t_k}, t_{k+1}) + \nabla_x c(\bar{x}_{t_k}, t_k)^T \Delta x = 0 \\
& \Delta x \geq -\bar{x}_{t_k}
\end{aligned}
$$

Theorem

- **A1: LGE is Strongly Regular at** $w_{t_k}^*$
- **A2:** $\bar{w}_{t_k}$ **Exists in Neighborhood and** $\exists\, \delta_r \geq 0$ s.t. $\|\bar{w}_{t_k} - w_{t_k}^*\| \leq L_\psi \|r(\bar{w}_{t_k}, t_k)\| \leq L_\psi \delta_r$

For sufficiently small Δt **,**

$$\|\bar{w}_{t_k} - w_{t_k}^*\| \leq L_\psi \delta_r \Rightarrow \|\bar{w}_{t_{k+1}} - w_{t_{k+1}}^*\| \leq L_\psi \delta_r$$

Analysis Straightforward Using Residual Bounds

Stability Holds Even if QP Solved to $O(\Delta t^2)$ **Accuracy**

2. Augmented Lagrangean Strategy: what if I am limited in memory and sometimes forced to terminate early even the QP resolution?

Augmented Lagrangean

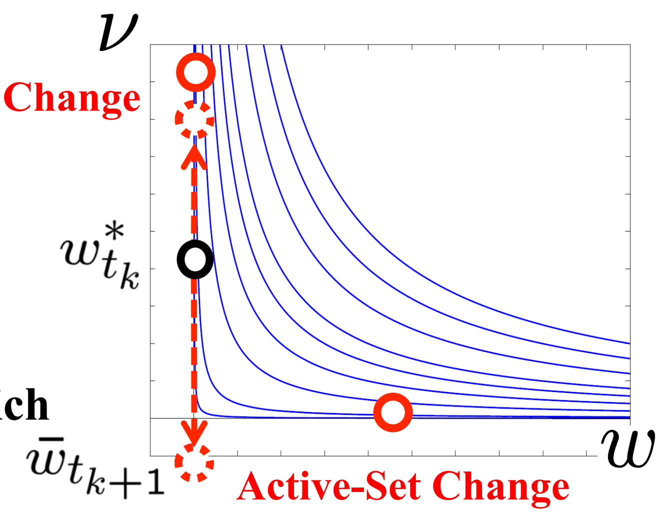
Iterative Linear Algebra to Solve QP

- Direct Linear Solvers Cannot be Terminated Early (Wasted Overhead)
- Complicated by Changing Active-Sets

Alternative: Barrier & Apply Iterative Solver to Indefinite KKT System (Smoothing)

$$\begin{aligned} \min_x \ & f(x,t) \\ \text{s.t. } & c(x,t) = 0 \\ & x \geq 0 \end{aligned} \quad \longleftrightarrow \quad \begin{aligned} \min \ & \phi(x,t) := f(x,t) - \mu \sum_{i=1}^{nx} \ln\left(x^{(i)}\right) \\ \text{s.t. } & c(x,t) = 0 \end{aligned}$$

$$\begin{bmatrix} \nabla_{xx}\mathcal{L}(\bar{w}_{t_k}, t_k) + \Sigma_{t_k} & \nabla_x c(\bar{x}_{t_k}, t_k) \\ \nabla_x c(\bar{x}_{t_k}, t_k)^T & \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \end{bmatrix} = - \begin{bmatrix} \nabla_x \phi(\bar{x}_{t_k}, t_{k+1}) + \nabla_x c(\bar{x}_{t_k}, t_{k+1}) \bar{\lambda}_{t_k} \\ c(\bar{x}_{t_k}, t_{k+1}) \end{bmatrix}$$



- Truncated Newton with PCG, QMR
- Barrier Linearization Leads to Large Errors
- Fast Indefinite Preconditioner Needed
- Plus, barrier introduces a large parameter which may severely affect stability

Augmented Lagrangean

Proposal: Augmented Lagrangean Penalty and Apply Projected Gauss-Seidel to QP

$$\begin{aligned} \min_x \;& f(x,t) \\ \text{s.t. } & c(x,t) = 0 \\ & x \geq 0 \end{aligned} \quad \longleftrightarrow \quad \begin{aligned} \min \;& \mathcal{L}_A(x,\lambda,t) := f(x,t) + \lambda^T c(x,t) + \frac{\rho}{2}\|c(x,t)\|^2 \\ \text{s.t. } & x \geq 0 \end{aligned}$$

$$\begin{aligned} \min \;& \nabla_x \mathcal{L}_A(\bar{x}_{t_k}, \bar{\lambda}_{t_k}, t_{k+1})^T \Delta x + \tfrac{1}{2}\Delta x^T \nabla_{xx}\mathcal{L}_A(\bar{x}_{t_k}, \bar{\lambda}_{t_k}, t_k)\Delta x \\ \text{s.t. } & \Delta x \geq -\bar{x}_{t_k} \end{aligned}$$

<u>Close to Manifold</u> Hessian of Augmented Lagrangean Remains at Least Positive Semi-Definite

Projected Gauss Seidel

$$\min_{w \geq \alpha} \quad \frac{1}{2} w^T \mathrm{M} w + \mathrm{b}^T w$$

For $k = 0, 1, \ldots, n_{iter}$

$$\begin{aligned} w_i^{k+1} &= -\frac{1}{\mathrm{M}_{ii}}\left(\mathrm{b}_i - \sum_{j<i} \mathrm{M}_{ij} w_j^{k+1} - \sum_{j>i} \mathrm{M}_{ij} w_j^k\right) \\ w_i^{k+1} &= \max\left(w_i^{k+1}, \alpha_i\right), \quad i = 1, \ldots, n \end{aligned}$$

- **Detects Multiple Active-Set Efficiently** *Morales et.al. 2008, Tasora et.al. 2009*
- **High Accuracy Requires Large Number of Iterations → Not if** Δt **Small! Ideal for us!**

Augmented Lagrangean

Algorithm:

Given $\bar{x}_{t_0}, \bar{\lambda}_{t_0}$, Δt, ρ, and n_{PGS},

1. Evaluate $\nabla_x \mathcal{L}_A(\bar{x}_{t_k}, \bar{\lambda}_{t_k}, t_{k+1}, \rho)$ and $\nabla_{xx} \mathcal{L}_A(\bar{x}_{t_k}, \bar{\lambda}_{t_k}, t_k, \rho)$.
2. Compute $\Delta \bar{x}_{t_{k+1}}$ applying n_{PGS} iterations to QP
3. Update $\bar{x}_{t_{k+1}} \leftarrow \bar{x}_{t_k} + \Delta \bar{x}_{t_{k+1}}$ and $\underbrace{\bar{\lambda}_{t_{k+1}} \leftarrow \bar{\lambda}_{t_k} + \rho\, c(\bar{x}_{t_{k+1}}, t_{k+1})}$.
4. $k \leftarrow k + 1$

First-Order Multiplier Update, *Hestenes 1969*

Avoids Major Operations

AugLag Penalty Acts as Parametric Perturbation of Lagrange Multipliers

Theorem

- **A1: Augmented Lagrangean - LGE is Strongly Regular at** $w^*_{t_k}$
- **A2:** $\bar{w}_{t_k}$ **Exists in Neighborhood and** $\exists\ \delta_r \geq 0$ s.t. $\|\bar{w}_{t_k} - w^*_{t_k}\| \leq L_\psi \|r(\bar{w}_{t_k}, t_k)\| \leq L_\psi \delta_r$

For sufficiently small Δt **and sufficiently large** ρ**,**

$$\|\bar{w}_{t_k} - w^*_{t_k}\| \leq L_\psi \delta_r \Rightarrow \|\bar{w}_{t_{k+1}} - w^*_{t_{k+1}}\| \leq L_\psi \delta_r$$

- **Conditions More Strict Due to Multiplier Error**
- **Tune** n_{PGS} **to Keep QP Solution Error** $O(\Delta t^2)$

Augmented Lagrangean

$$\min_{w \geq \alpha} \quad \frac{1}{2} w^T \mathrm{M} w + \mathrm{b}^T w$$

For $k = 0, 1, ..., n_{iter}$

$$w_i^{k+1} = -\frac{1}{\mathrm{M}_{ii}} \left(\mathrm{b}_i - \sum_{j<i} \mathrm{M}_{ij} w_j^{k+1} - \sum_{j>i} \mathrm{M}_{ij} w_j^k \right)$$

$$w_i^{k+1} = \max \left(w_i^{k+1}, \alpha_i \right), \quad i = 1, ..., n$$

Remarks:

1. **Projected GS is Powerful Paradigm for <u>Linear MPC</u>**

   - Fixed Matrix, Block Parallelizable (Multi-Thread)

2. **Even if Dynamic System is SLOW….**

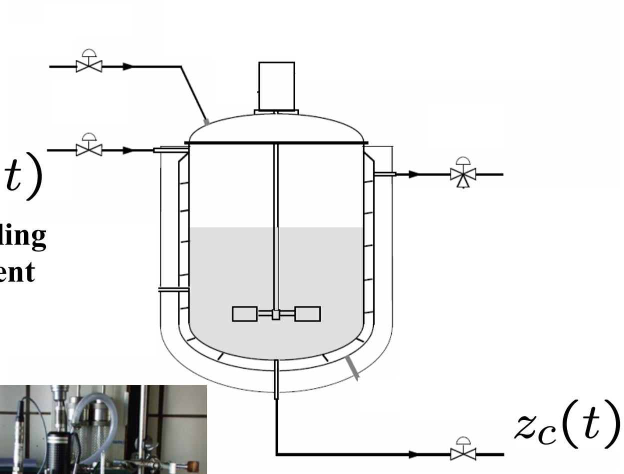
   - Solve QP at High Frequency (Open-Loop) to Keep Track of Solution Manifold

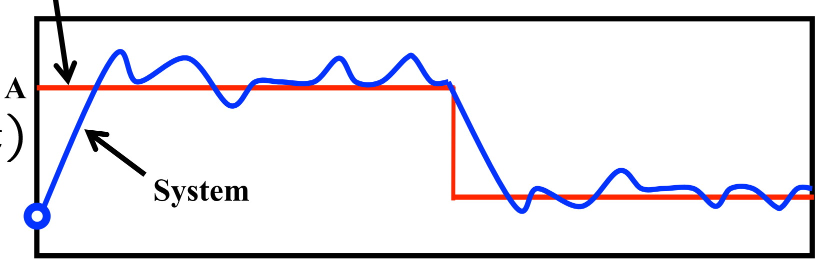
   - Once Control is Needed, the Solution is <u>Very Close</u>, use as Warm-Start

3. Numerical Case Study

Numerical Case Study

Control of Polymerization Reactor





Converted to NLP by Applying Implicit Euler Scheme

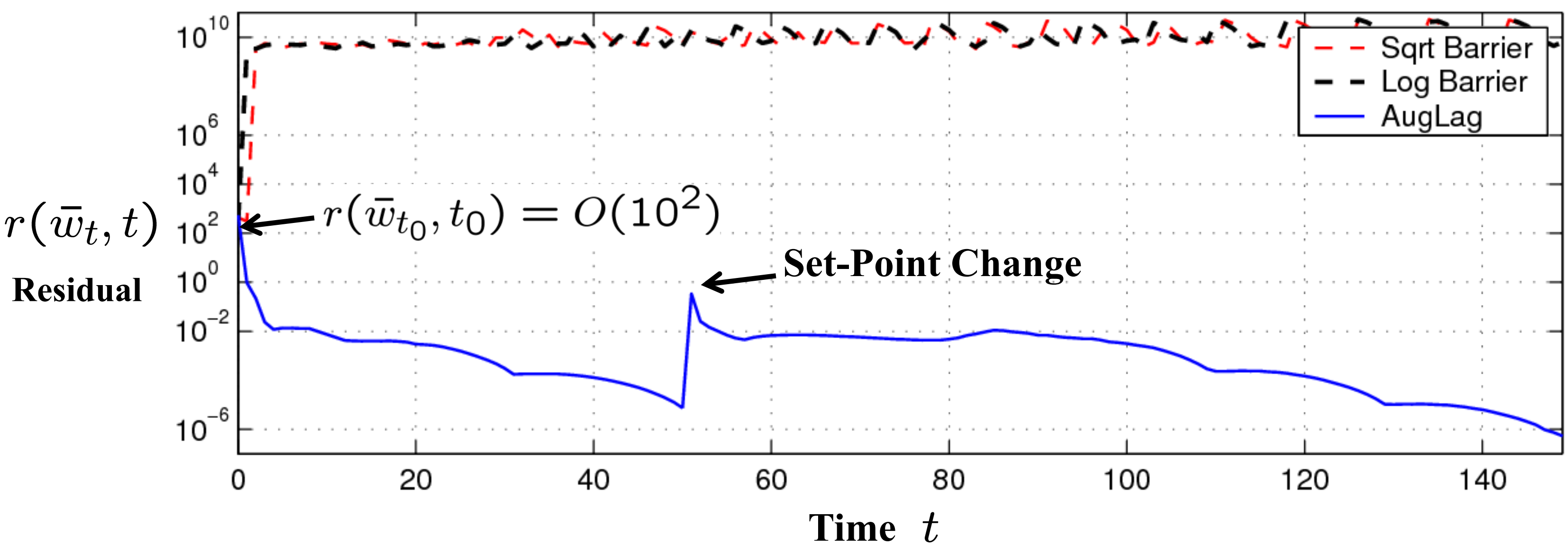
Numerical Case Study

Numerical Tests

- Comparison Against Barrier Smoothing *Heath, 2004, Ohtsuka, 2004*

1) $\mu \cdot \log(x - x^{min}) + \mu \cdot \log(x^{max} - x)$ **2)** $\mu \cdot \text{sqrt}(x - x^{min}) + \mu \cdot \text{sqrt}(x^{max} - x)$

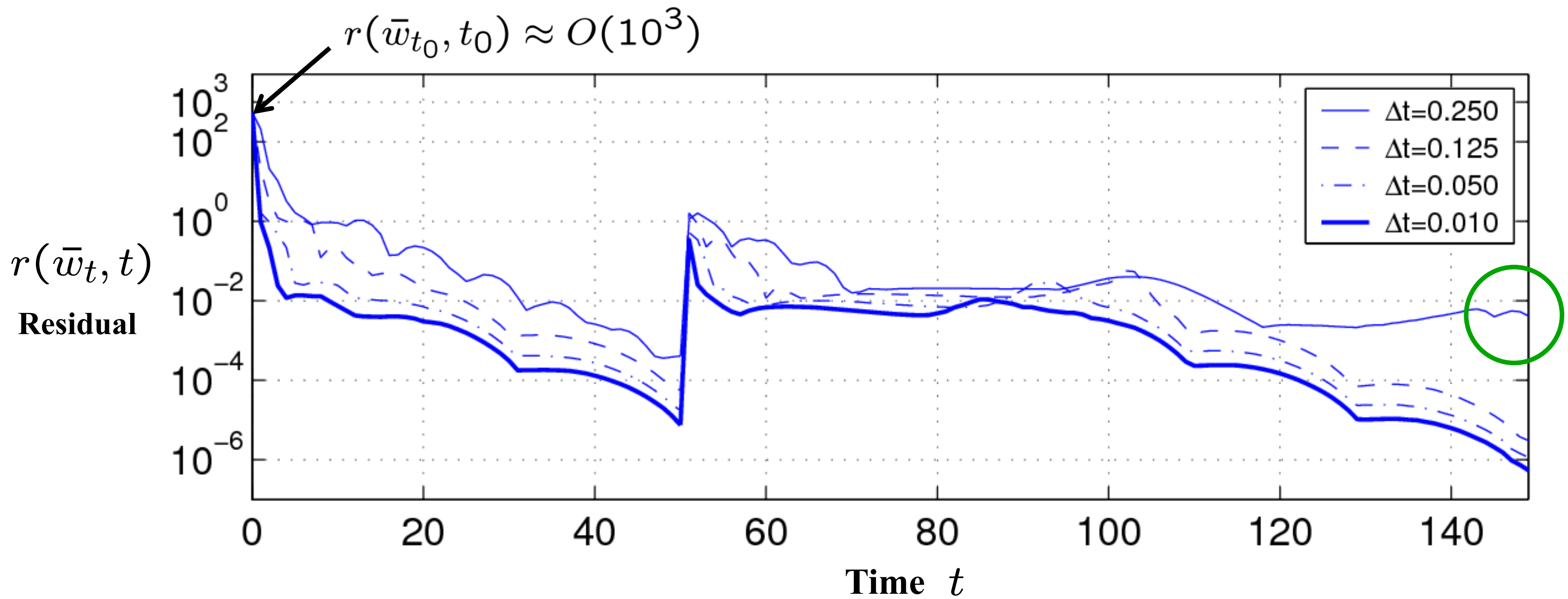
\- $n_{PGS} = 25,\ \Delta t = 0.025,\ \rho = 100$



Smoothing is Numerically <u>Unstable</u> – Active-Set Changes
Augmented Lagrangean Stands Relatively Large Initial Errors

Numerical Case Study

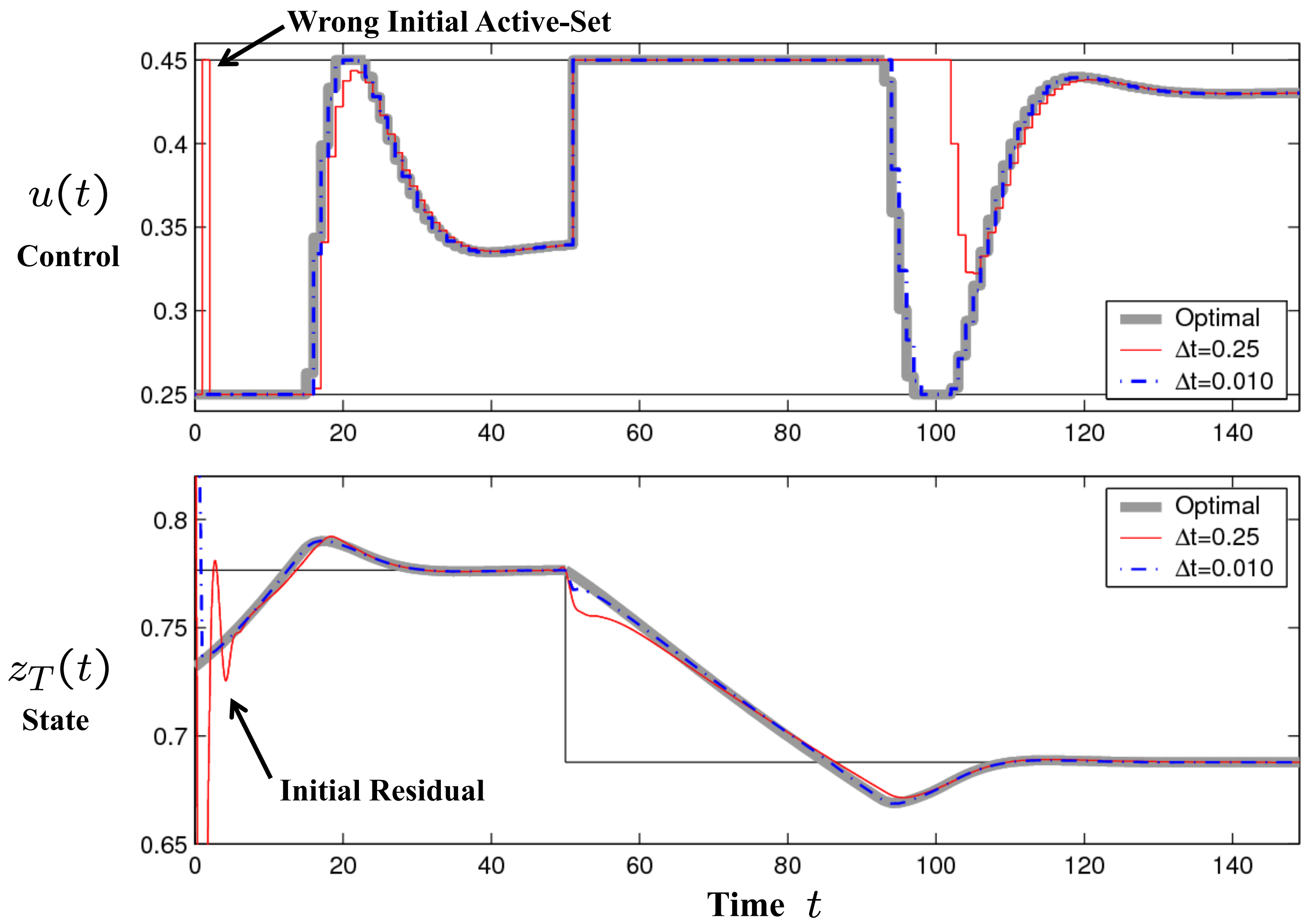
Effect of Time Step Δt



Sampling Time Restricted by Time Needed to Perform n_{PGS} Iterations

Numerical Case Study

Optimal vs. Approximate Profiles



4. Conclusions and Future Work

Conclusions and Future Work

Motivation

- Accurate Search Steps Not Necessarily Best in Time-Critical Environments
- Cheap Strategies to Ensure $\Delta t \rightarrow 0$ and Still Guarantee Error Stability

Generalized Equations

- Powerful Framework for Analysis of On-Line NLP Strategies

Augmented Lagrangean Strategy

- Projected Gauss-Seidel for High-Frequency QP Solutions

Work Needed

- Convergence
- Time-Adaptive Schemes
- Multi-Thread Implementations, Industrial Examples
- Avoid Augmented Lagrangean -- Projection Methods for General QPs

On-Line Nonlinear Programming as a Generalized Equation

Victor M. Zavala

Postdoctoral Researcher
Mathematics and Computer Science Division
Argonne National Laboratory

vzavala@mcs.anl.gov

Joint Work with: Mihai Anitescu

TWCCC Fall Meeting
September, 2009